Questions (and Answers) on the Semantic Web

Oslo, Norway, 2006-09-20

Ivan Herman, W3C

We all know that, right?

- The Semantic Web Artificial Intelligence on the Web
- It relies on centrally controlled ontologies for “meaning”
  - *as opposed to a democratic, bottom–up control of terms*
- One has to add metadata to all Web pages, convert all relational databases, and XML data to use the Semantic Web
- It is just an ugly application of XML
- One has to learn formal logic, knowledge representation techniques, description logic, etc
- It is, essentially, an academic project, of no interest for industry
- …

WRONG!!!!

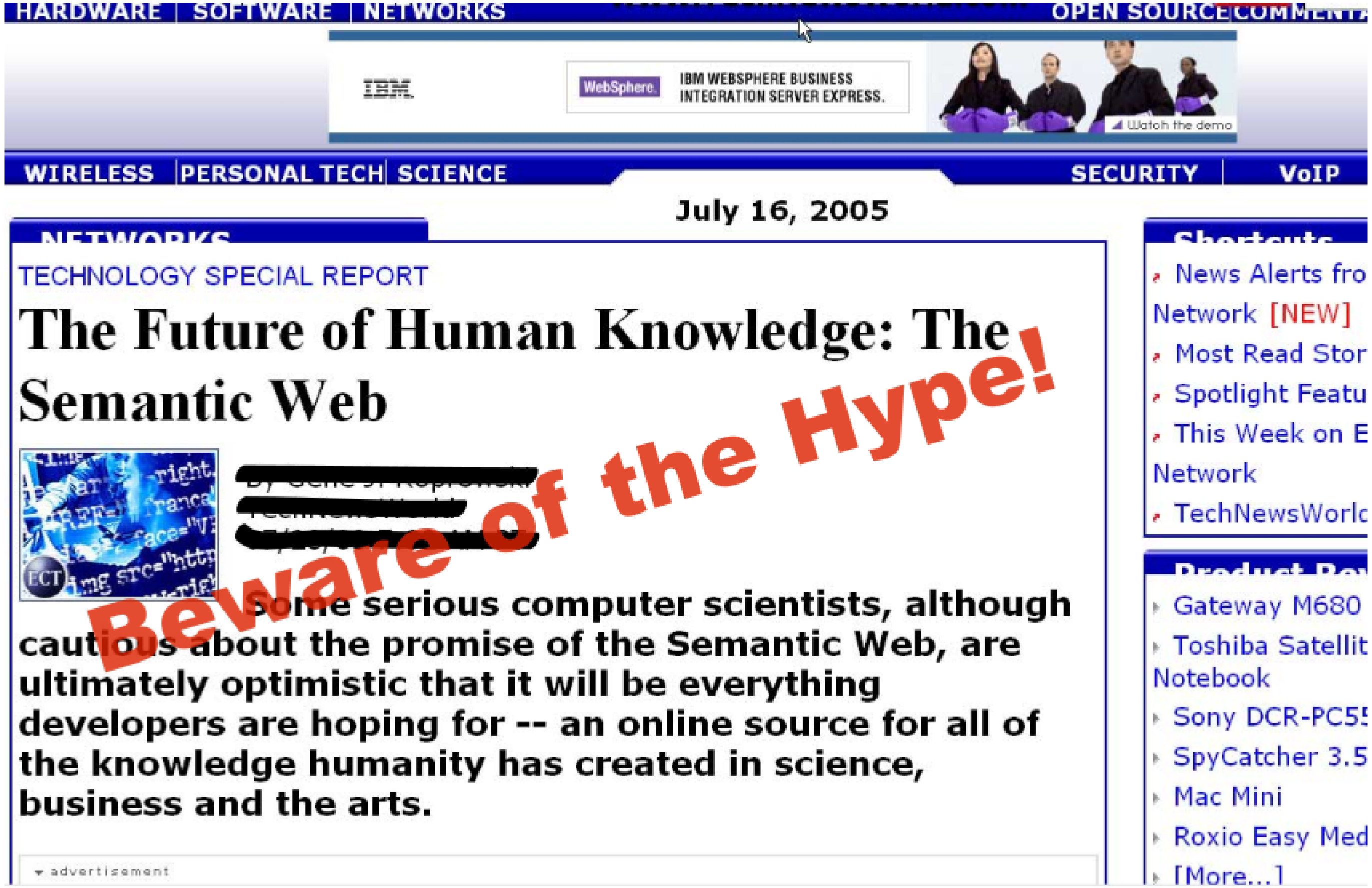
- ~~The Semantic Web Artificial Intelligence on the Web~~
- ~~It relies on centrally controlled ontologies for “meaning”~~
  - *~~as opposed to a democratic, bottom–up control of terms~~*
- ~~One has to add metadata to all Web pages, convert all relational databases, and XML data to use the Semantic Web~~
- ~~It is just an ugly application of XML~~
- ~~One has to learn formal logic, knowledge representation techniques, description logic, etc~~
- ~~It is, essentially, an academic project, of no interest for industry~~
- ~~…~~

Goal of this presentation…

- There are lots of myths around the Semantic Web
- This presentation will try to de-mystify at least some of those…

Is the Semantic Web AI on the Web?

No!



July 16, 2005

TECHNOLOGY SPECIAL REPORT

The Future of Human Knowledge: The Semantic Web

[redacted]

Some serious computer scientists, although cautious about the promise of the Semantic Web, are ultimately optimistic that it will be everything developers are hoping for -- an online source for all of the knowledge humanity has created in science, business and the arts.

So what *is* the Semantic Web?

- Humans can easily “connect the dots” when browsing the Web…
  - *you disregard advertisements*
  - *you “know” (from the context) that this link is interesting and goes to my CV; whereas the that one is without interest*
  - *etc.*
- … but machines can’t!
- The goal is to have a *Web of Data* to ensure smooth integration with data, too
- Let us see just some application examples…

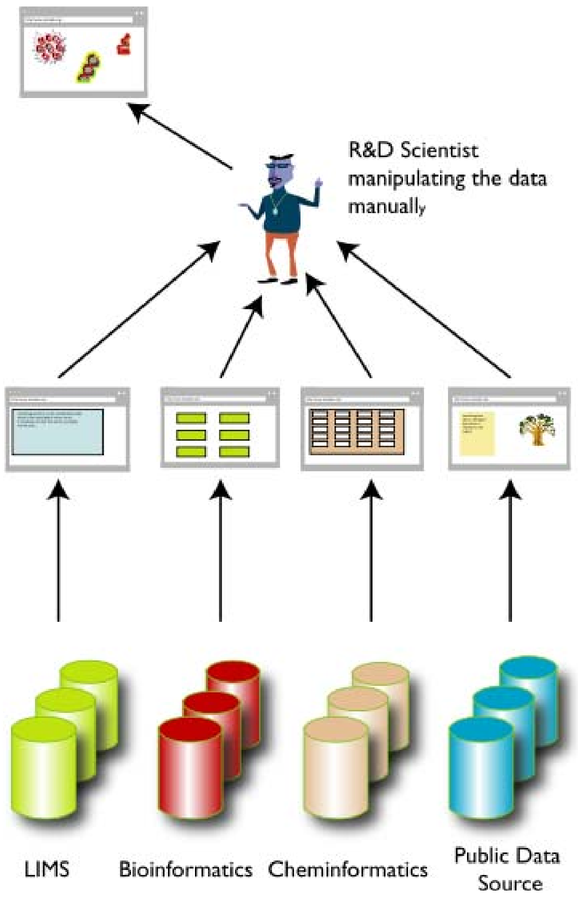
Example: Automatic Airline Reservation

- Your automatic airline reservation
  - *knows about your preferences*
  - *builds up knowledge base using your past*
  - *can combine the local knowledge with remote services:*
    - airline preferences
    - dietary requirements
    - calendaring
    - etc
- It communicates with *remote* information (i.e., on the Web!)
- (M. Dertouzos: The Unfinished Revolution)

Example: data(base) integration

- Databases are very different in structure, in content
- Lots of applications require managing *several* databases
  - *after company mergers*
  - *combination of administrative data for e-Government*
  - *biochemical, genetic, pharmaceutical research*
  - *etc.*
- Most of these data are now on the Web (though not necessarily public yet)

Example: data integration in life sciences



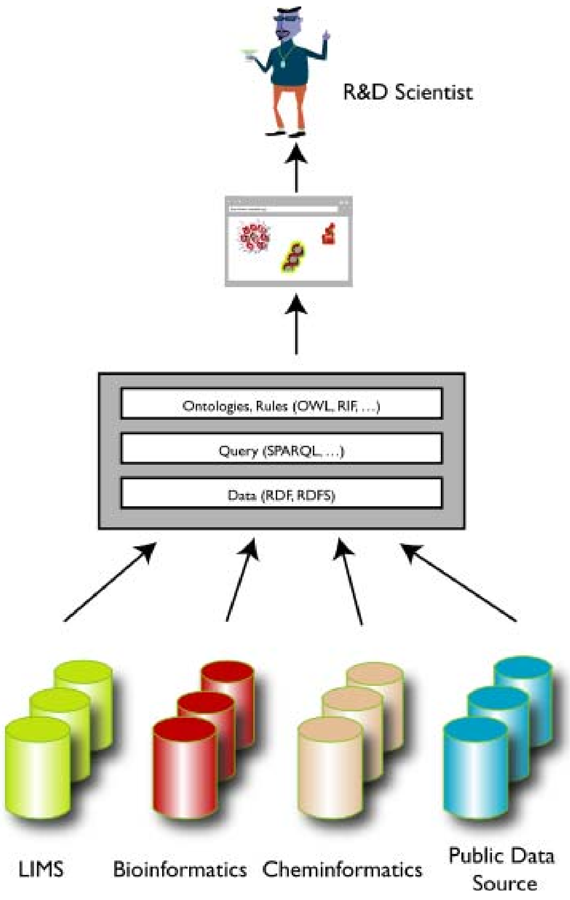
R&D Scientist manipulating the data manually

LIMS

Bioinformatics

Cheminformatics

Public Data Source



R&D Scientist

Ontologies, Rules (OWL, RIF, …)

Query (SPARQL, …)

Data (RDF, RDFS)

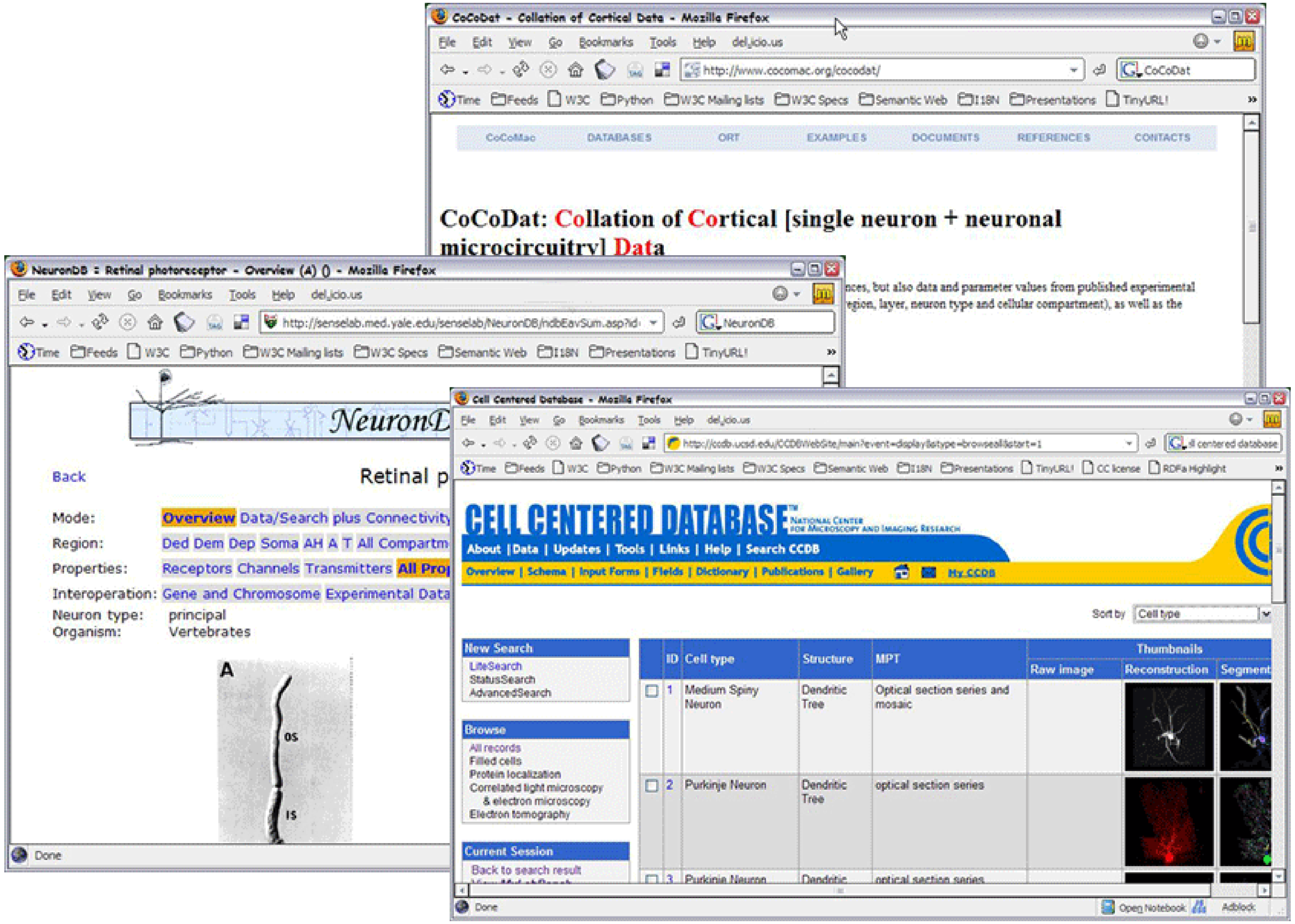
LIMS

Bioinformatics

Cheminformatics

Public Data Source

And the problem *is* real



CoCoDat - Collation of Cortical Data - Mozilla Firefox
File Edit View Go Bookmarks Tools Help del.icio.us
http://www.cocomac.org/cocodat/
CoCoDat
Time Feeds W3C Python W3C Mailing lists W3C Specs Semantic Web I18N Presentations TinyURL!
CoCoMac DATABASES ORT EXAMPLES DOCUMENTS REFERENCES CONTACTS

CoCoDat: Collation of Cortical [single neuron + neuronal microcircuitry] Data

nces, but also data and parameter values from published experimental
region, layer, neuron type and cellular compartment), as well as the

NeuronDB : Retinal photoreceptor - Overview (A) () - Mozilla Firefox
File Edit View Go Bookmarks Tools Help del.icio.us
http://senselab.med.yale.edu/senselab/NeuronDB/ndbEavSum.asp?id
NeuronDB
Time Feeds W3C Python W3C Mailing lists W3C Specs Semantic Web I18N Presentations TinyURL!
NeuronD
Back

Retinal p

Mode: Overview Data/Search plus Connectivity
Region: Ded Dem Dep Soma AH A T All Compartm
Properties: Receptors Channels Transmitters All Prop
Interoperation: Gene and Chromosome Experimental Data
Neuron type: principal
Organism: Vertebrates

A
OS
IS
Done
Cell Centered Database - Mozilla Firefox
File Edit View Go Bookmarks Tools Help del.icio.us
http://ccdb.ucsd.edu/CCDBWebSite/main?event=display&type=browseall&start=1
ll centered database
Time Feeds W3C Python W3C Mailing lists W3C Specs Semantic Web I18N Presentations TinyURL! CC license RDFa Highlight
CELL CENTERED DATABASE™ National Center for Microscopy and Imaging Research
About | Data | Updates | Tools | Links | Help | Search CCDB
Overview | Schema | Input Forms | Fields | Dictionary | Publications | Gallery
My CCDB
Sort by Cell type
New Search
LiteSearch
StatusSearch
AdvancedSearch
Browse
All records
Filled cells
Protein localization
Correlated light microscopy & electron microscopy
Electron tomography
Current Session
Back to search result
Done
Open Notebook
Adblock

	ID	Cell type	Structure	MPT	Thumbnails		
					Raw image	Reconstruction	Segment
□	1	Medium Spiny Neuron	Dendritic Tree	Optical section series and mosaic			
□	2	Purkinje Neuron	Dendritic Tree	optical section series			
□	3	Purkinje Neuron	Dendritic	optical section series			

So what *is* the Semantic Web?

The Semantic Web is… the Web of Data

- It allows machines to “connect the dots”
- It provides a common framework to share data on the Web across application boundaries

And what is the relationship to AI?

- *Some* technologies in the Semantic Web has benefited from AI research and development (see later)
- Semantic Web has also brought some new concerns, problems, use cases to AI
- But AI has many many different problems that are not related to the Web at all (image understanding is a good example)

A possible comparison

Smarter machines

- teach computers to infer the meaning of Web data
  - *natural language, image recognition, etc.*
- …this is the Artificial Intelligence approach

Smarter data

- Make data easier for machines to find, access and process
  - *express data and meaning in standard machine-readable format*
  - *support decentralized definition and management, across the network*
- …this is the Semantic Web approach

All right, but what is RDF then?

RDF

- For all applications listed above the issues are *to create relations among resources on the Web* and to *interchange those data*
- Pretty much like (hyper)links on the traditional web, except that:
  - *there is no notion of “current” document; ie, relationship is between any two resources*
  - *a relationship must have a name: a link to my CV should be differentiated from a link to my Calendar*
  - *there is no attached user-interface action like for a hyperlink*

RDF (cont.)

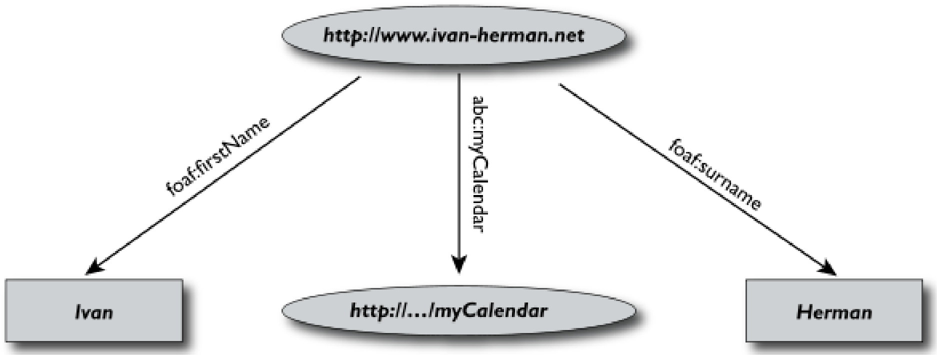
- *RDF is a model for such relationships and interchange*
  - *to be a bit more techie: it is a model of* `(s p o)` *triplets with* `p` *naming the relationship between* `s` *and* `o`
- URI-s are used as universal *naming* tools, including for properties (after all, “U” stands for “Universal”…)
- That is it (essentially)! Nothing very complex…

But isn’t RDF simply an (ugly) XML application?

RDF is a graph!

- As we already said: RDF is a set of relationships
- An `(s,p,o)` triple can be viewed as a labeled edge in a graph
  - *i.e., a set of RDF statements is a directed, labeled graph*
    - the nodes represent the resources that are bound
    - the labeled edges are the relationships with their names
- This set must be serialized for machines; this can be done into XML (using RDF/XML), or to other formats (Turtle, N-Triples, TriX, …)
- Think in terms of graphs, the rest is syntactic sugar!

A Simple RDF Example



```
<rdf:Description rdf:about="http://www.ivan-herman.net">
    <foaf:name>Ivan</foaf:name>
    <abc:myCalendar rdf:resource="http://…/myCalendar"/>
    <foaf:surname>Herman</foaf:surname>
</rdf:Description>
```

Yes, RDF/XML has its Problems

- RDF/XML was developed in the “prehistory” of XML
  - *e.g., even namespaces did not exist!*
- Coordination was not perfect, leading to problems
  - *the syntax cannot be checked with XML DTD-s*
  - *XML Schemas are also a problem*
  - *encoding is verbose and complex (simplifications lead to confusions…)*
- but there is too much legacy code to change it

Use, e.g., Turtle if you prefer…

```
<http://www.ivan-herman.net>
  foaf:firstName "Ivan";
  abc:myCalendar <http://.../myCalendar>;
  foaf:surname "Herman".
```

- Again: these are all just syntactic sugar!
- RDF environments often understand several serialization syntaxes
- In some cases, authoring tools hide the details anyway!

But what has RDF to do with data integration?

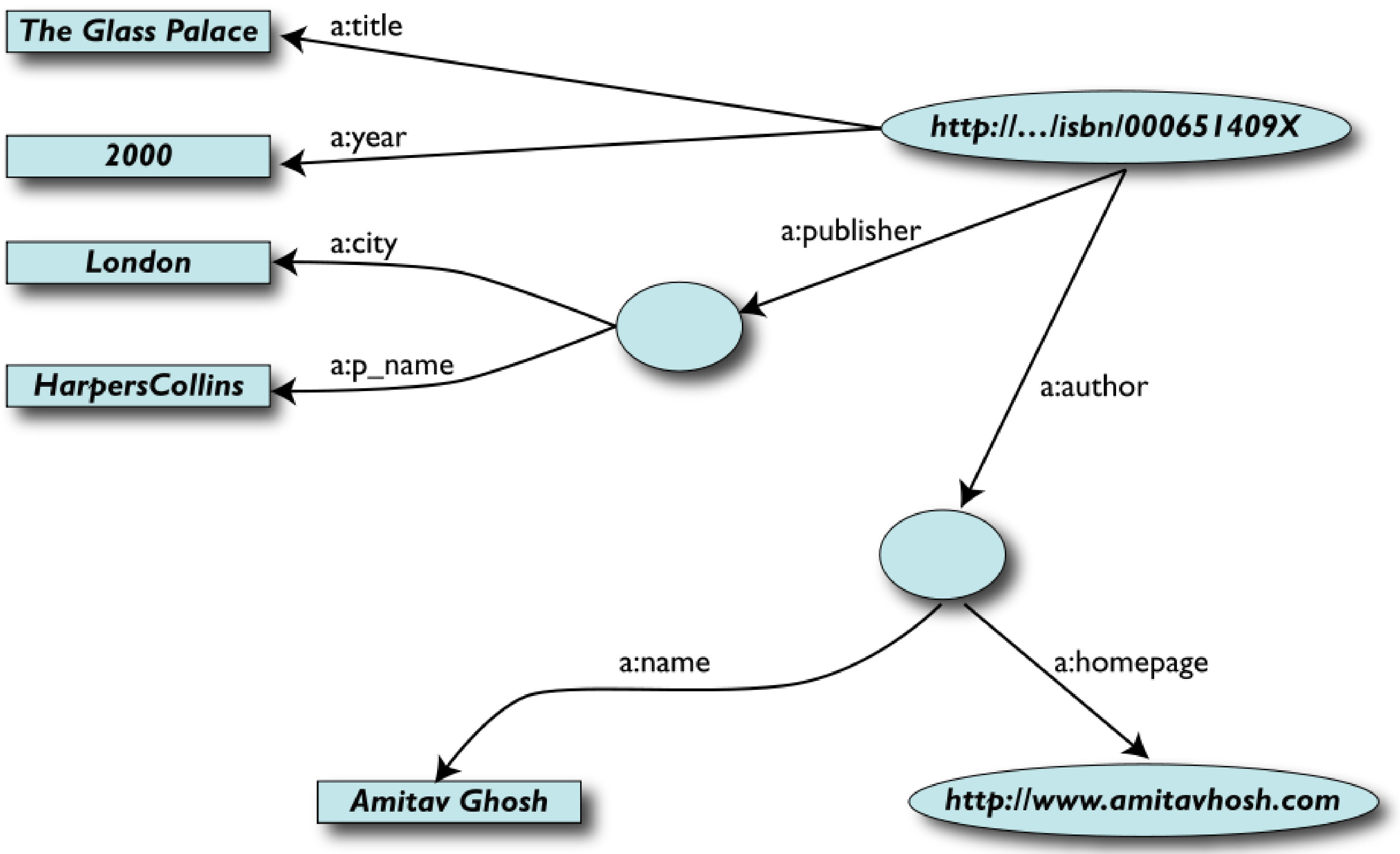
Consider this (simplified) bookstore data set

ID	Author	Title	Publisher	Year
ISBN 0-00-651409-X	id_xyz	The Glass Palace	id_qpr	2000

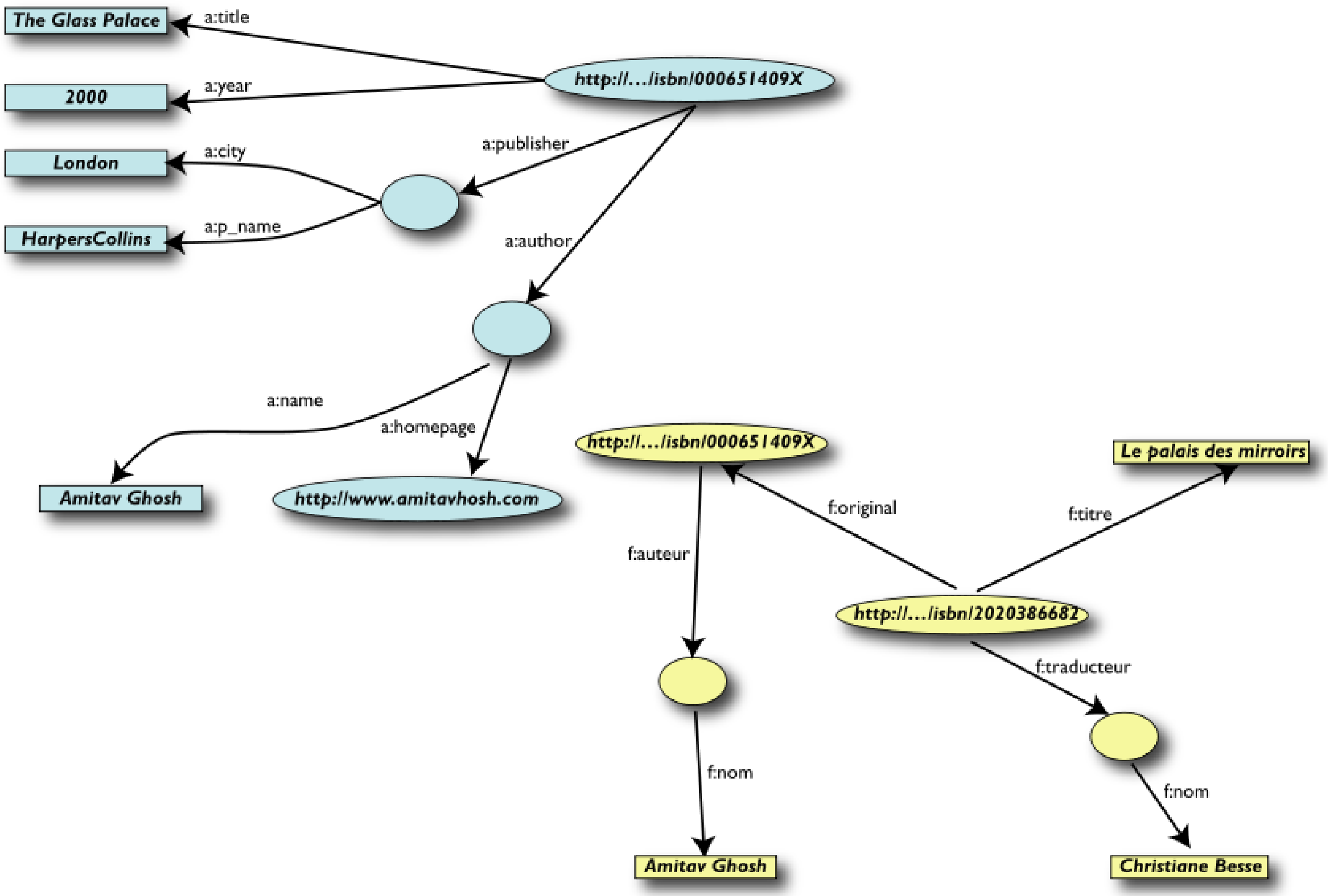
ID	Name	Home page
id_xyz	Amitav Ghosh	http://www.amitavghosh.com/

ID	Publisher Name	City
id_qpr	Harper Collins	London

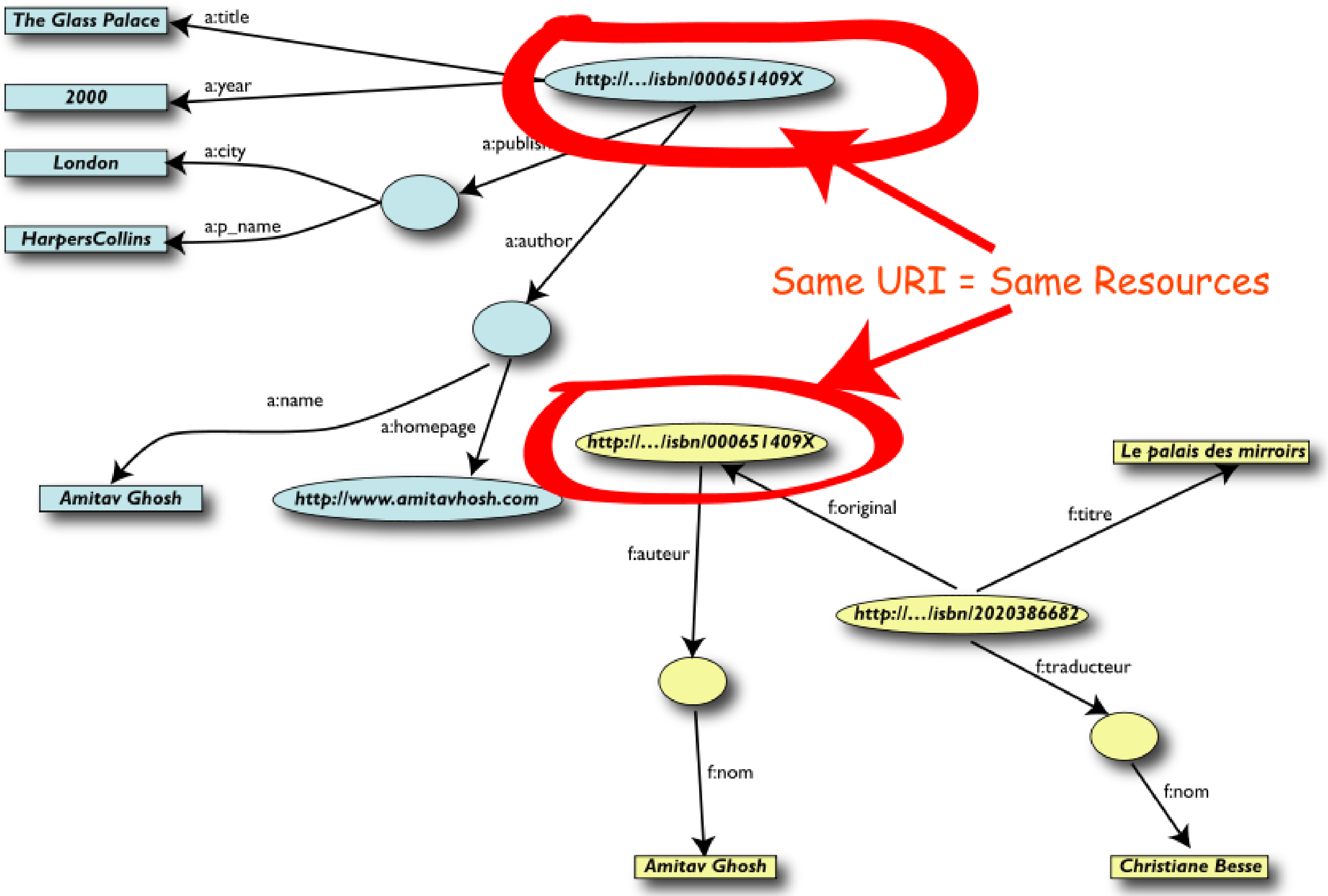
Export your data as a set of relations…



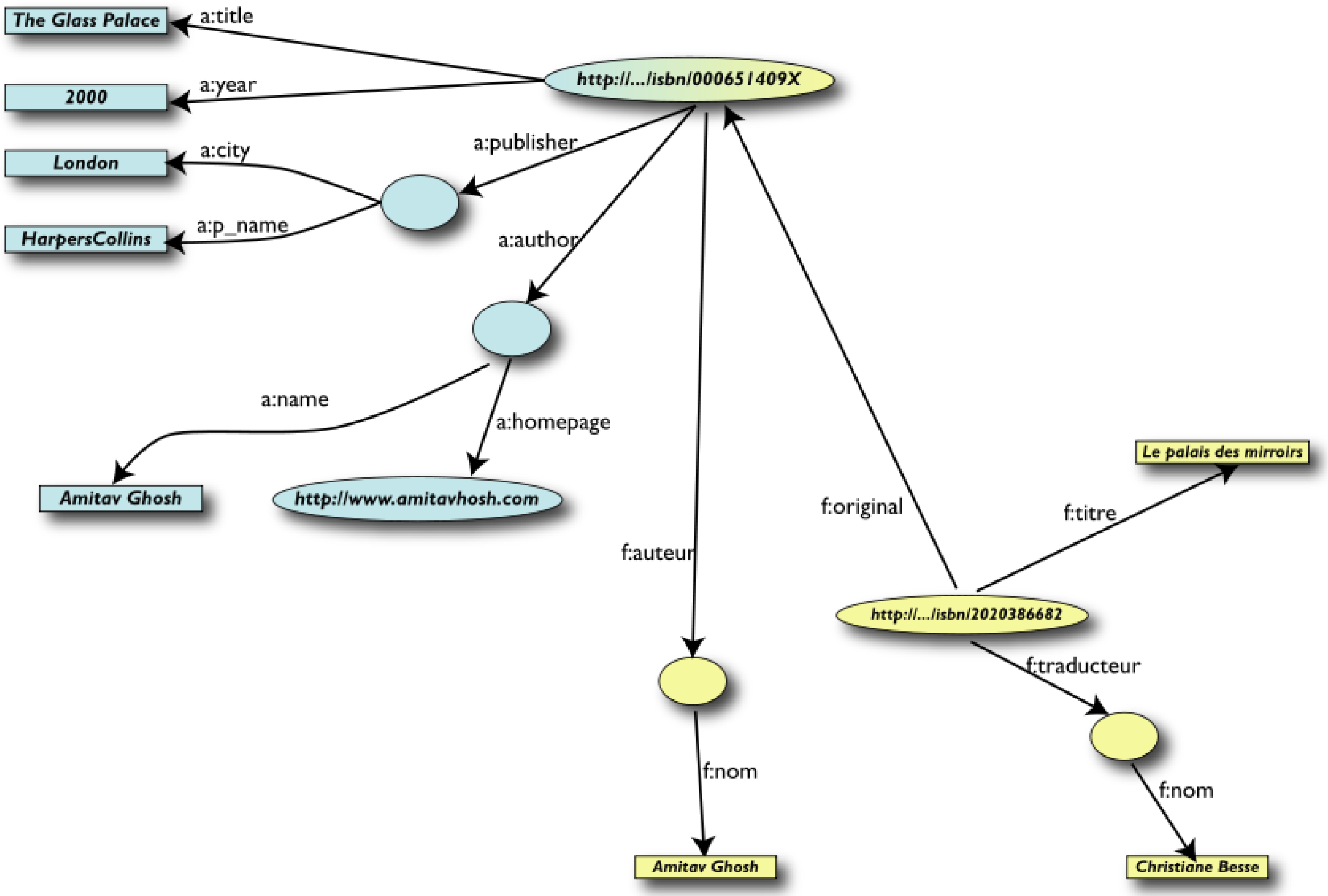
Add the data from another publisher…



Start merging…



Simple integration…



Note the role of URI-s!

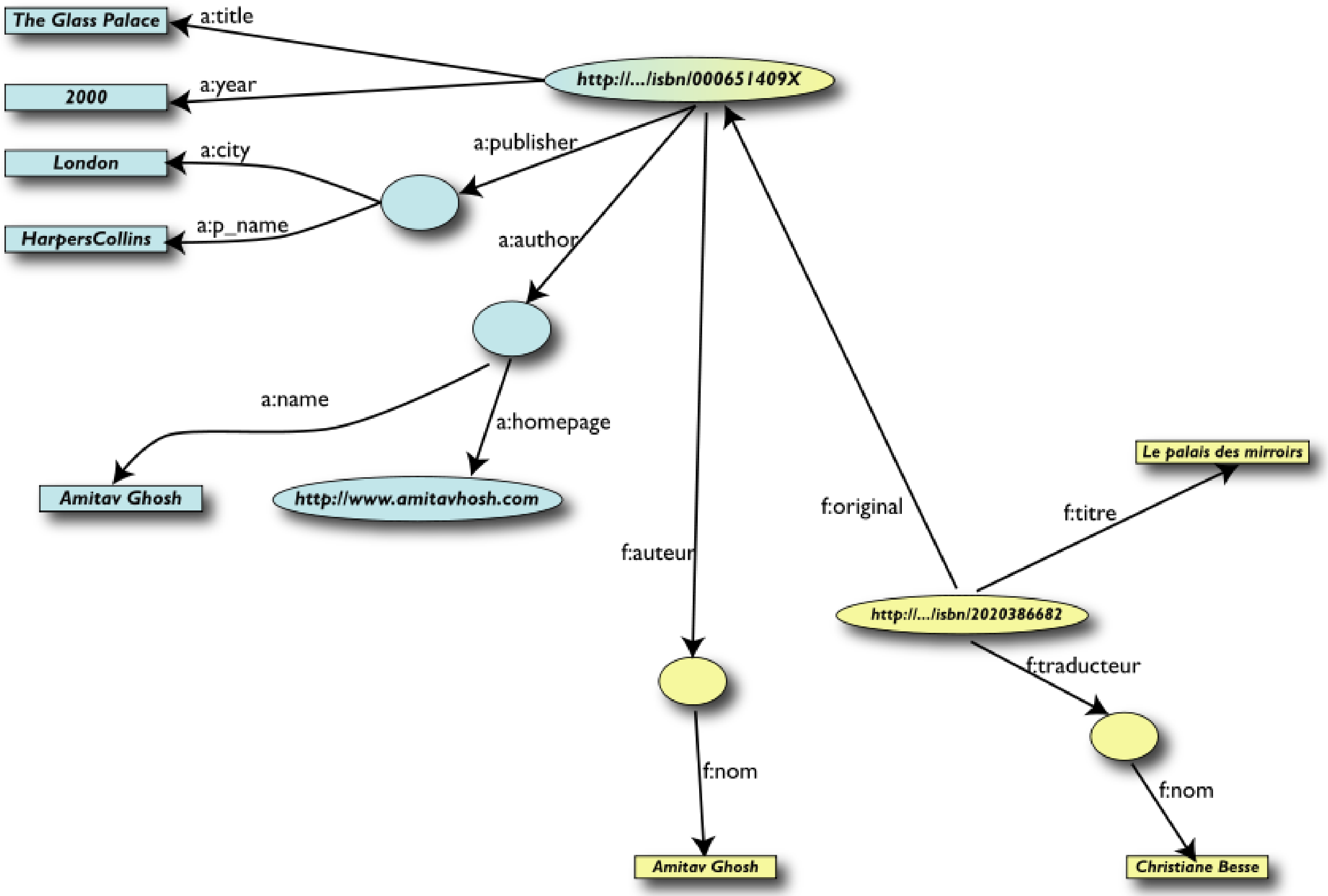
- The URI-s made the merge possible
- URI-s ground RDF into the Web
- URI-s make this the Semantic *Web*

So what is then the role of ontologies and/or rules?

A possible short answer

- *Ontologies/rules are there to help integration*
- Let us come back to our example…

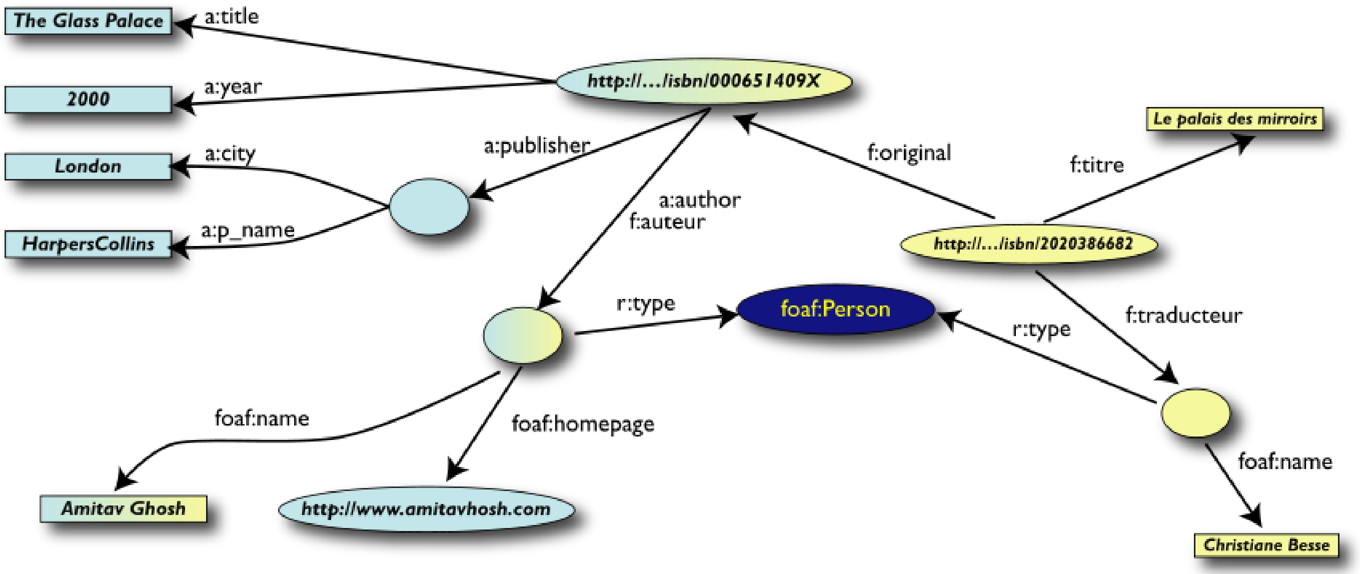
This is where we are…



Our merge is not complete yet…

- We “feel” that `a:author` and `f:auteur` should be the same
- But an automatic merge doest not know that!
- Let us add some extra information to the merged data:
  - *`a:author` same as `f:auteur`*
  - *both identify a “Person”:*
    - a term that a community has already defined (part of the “FOAF” terminology)
    - a “Person” is uniquely identified by his/her name and, say, homepage
    - it can be used as a “category” for certain type of resources
    - we can also identify, say, `a:name` with `foaf:name`

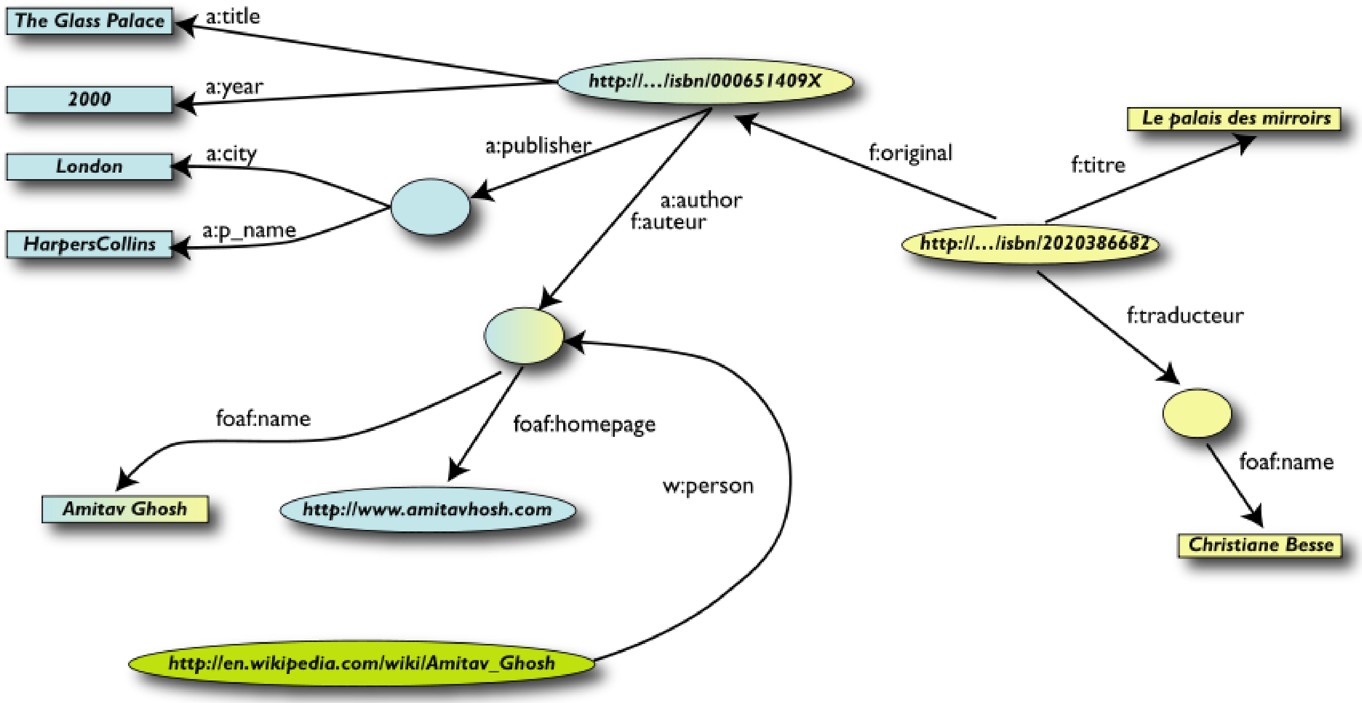
Better merge: richer queries are possible!



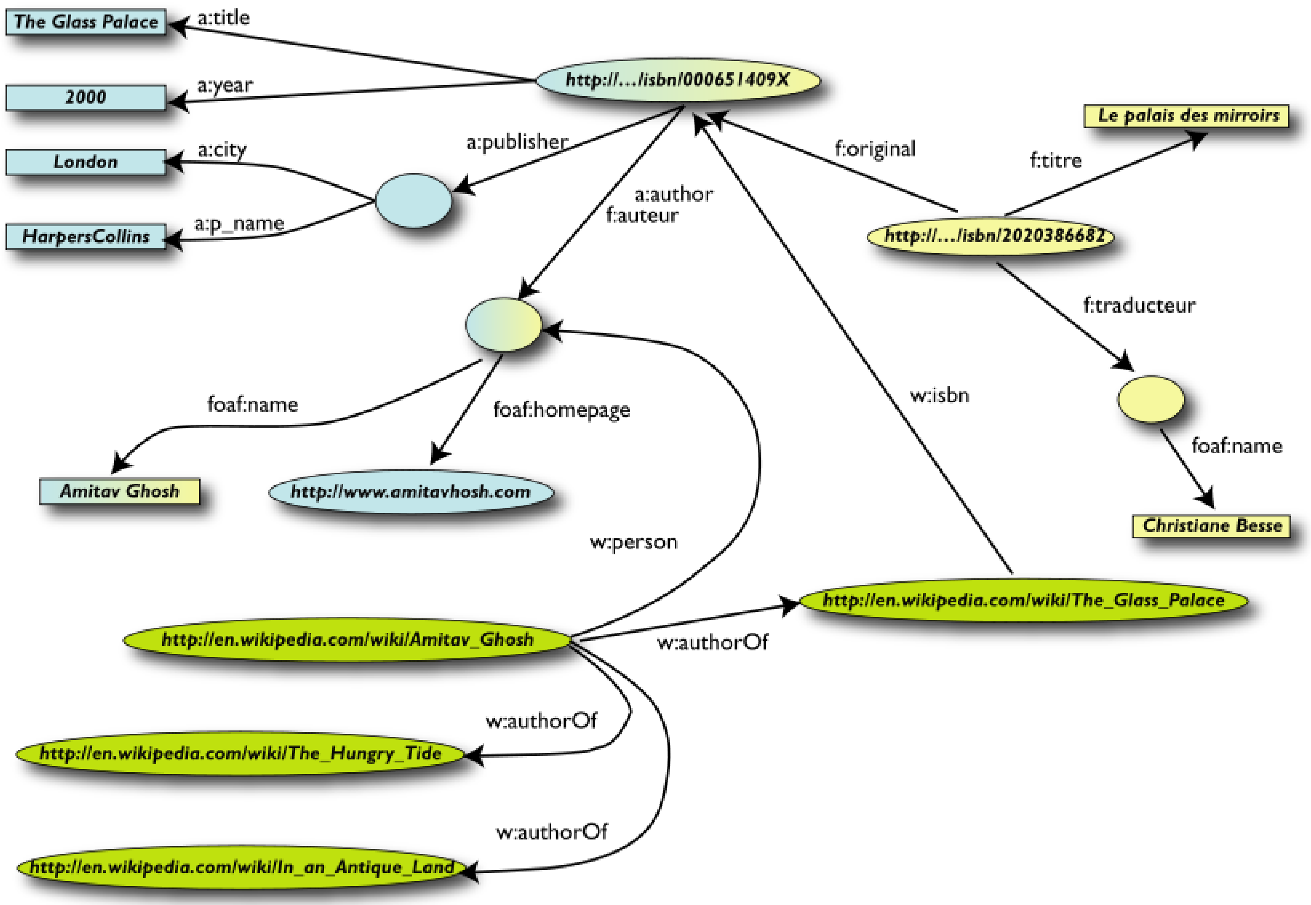
What we did: we used ontologies…

- We said:
  - *`a:author` same as `f:auteur`*
  - *both identify a "Person":*
    - *a term that a community has already defined*
    - *a "Person" is uniquely identified by his/her name and, say, homepage*
    - *it can be used as a "category" for certain type of resources*
    - *we can also identify, say, `a:name` with `foaf:name`*
- These statements can be described in an ontology (or, alternatively, with rules)
- The ontology/rule serves as some sort of a "glue"

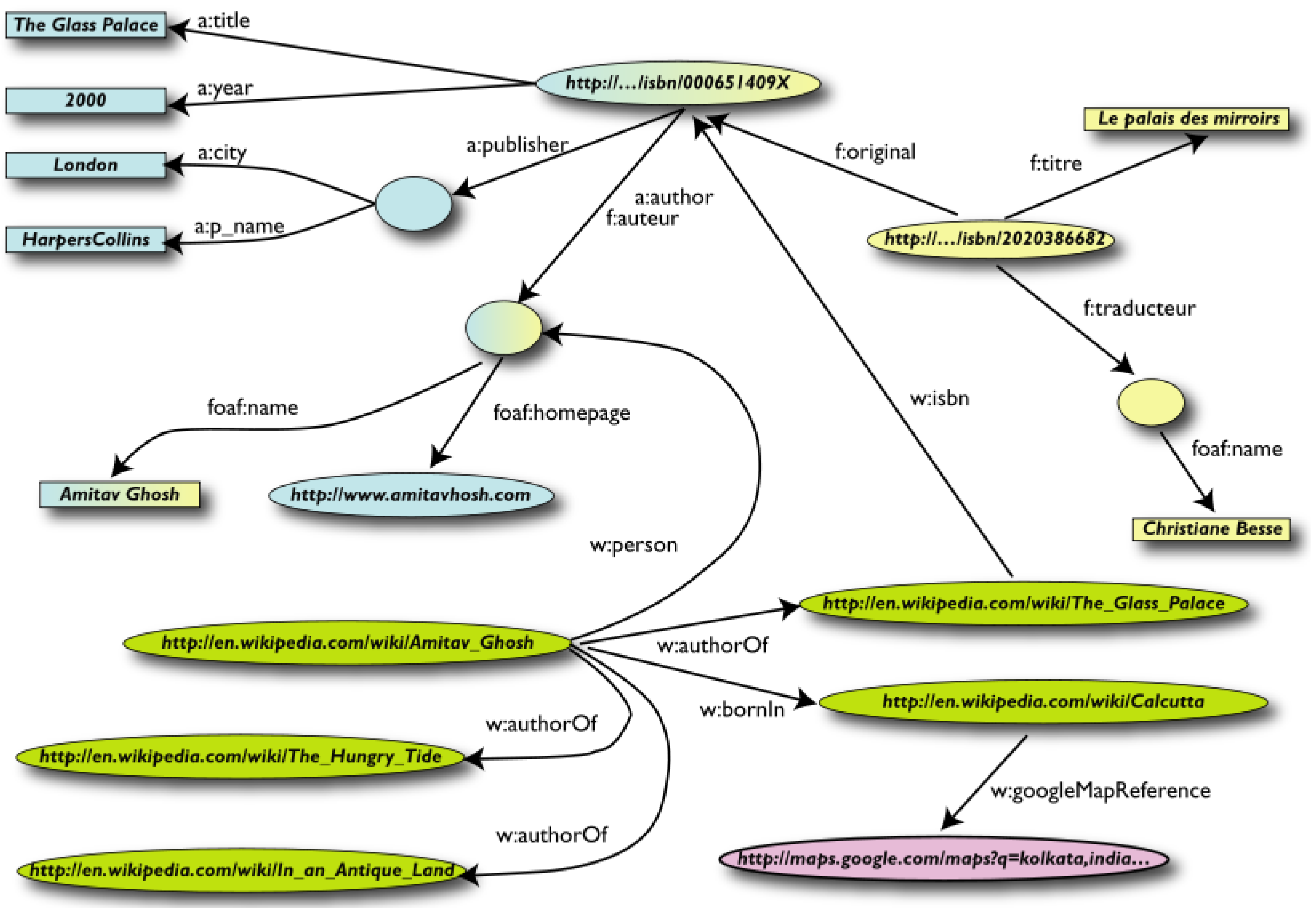
And then the merge may go on…



…and on…


The Glass Palace
a:title
2000
a:year
http://…/isbn/000651409X
London
a:city
a:publisher
HarpersCollins
a:p_name
a:author
f:auteur
f:original
Le palais des mirroirs
f:titre
http://…/isbn/2020386682
f:traducteur
foaf:name
foaf:homepage
Amitav Ghosh
http://www.amitavhosh.com
w:isbn
foaf:name
Christiane Besse
w:person
http://en.wikipedia.com/wiki/The_Glass_Palace
http://en.wikipedia.com/wiki/Amitav_Ghosh
w:authorOf
w:authorOf
http://en.wikipedia.com/wiki/The_Hungry_Tide
w:authorOf
http://en.wikipedia.com/wiki/In_an_Antique_Land

…and on…


The Glass Palace
a:title
2000
a:year
http://…/isbn/000651409X
London
a:city
HarpersCollins
a:p_name
a:publisher
a:author
f:auteur
f:original
Le palais des mirroirs
f:titre
http://…/isbn/2020386682
f:traducteur
foaf:name
Christiane Besse
foaf:name
Amitav Ghosh
foaf:homepage
http://www.amitavhosh.com
w:isbn
w:person
http://en.wikipedia.com/wiki/Amitav_Ghosh
http://en.wikipedia.com/wiki/The_Glass_Palace
w:authorOf
w:bornIn
http://en.wikipedia.com/wiki/Calcutta
w:authorOf
http://en.wikipedia.com/wiki/The_Hungry_Tide
w:googleMapReference
w:authorOf
http://en.wikipedia.com/wiki/In_an_Antique_Land
http://maps.google.com/maps?q=kolkata,india…

Is that surprising?

- Maybe but, in fact, no…
- What happened via automatic means is done all the time by the (human) users of the Web!
- The difference: a bit of extra rigor (eg, *naming* the relationships), extra information (eg, identifying relationships) and machines could do this, too

A very important issue: “schema independence”

- The queries (ie, the application) sees the RDF data only (with references to “real” data)
- If the structure (“schema”) of the database changes, only the mapping to RDF has to be changed
  - *this is a very local change*
- Ie, the RDF layer is very robust vis-a-vis schema evolution (not only to schema differences)

You remember this statement?

- *It relies on giant, centrally controlled ontologies for “meaning”*

- Ontologies are usually developed by communities and they are to be *shared*
  - *in fact, in our example, we used an ontology called “FOAF”*

And this?

- *One has to learn formal logic, knowledge representation techniques, description logic, etc, to understand the Semantic Web and be able to use it*

- This “glue” does *not* have to be complex, it may be of a few lines only
  - *“a little semantics can take you far…”*

So what does “inference” means on the Semantic Web?

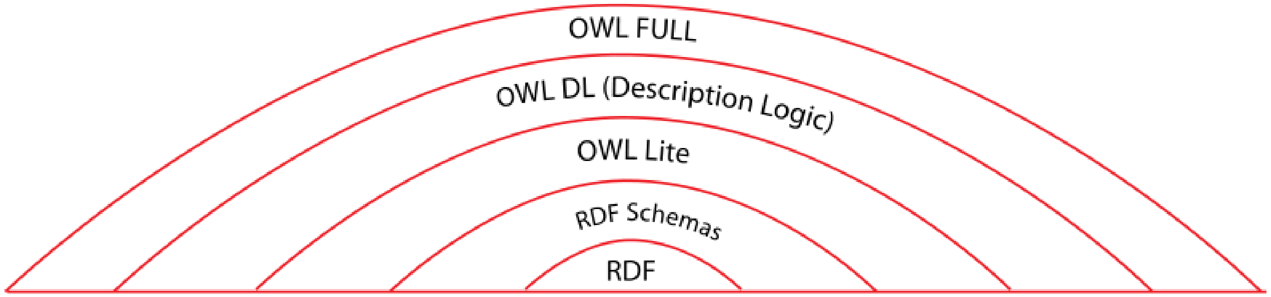
How do you “deduce” things?

Remember the “same as”?

- We said: `a:name` same as `foaf:name`
- What this meant, in ontology terms:
  - *if* `(P a:name L)` *is present, then* `(P foaf:name L)` *should be present, too (and vice versa)*
- But what this also means is that:
  - *whereas* `(P a:name L)` *is in the original data*
  - `(P foaf:name L)` *is an added (or “deduced”) relationship by virtue of the ontology we used*
- Ie: *“inference” means discovery of new relationships!*

Tradeoffs

- What can be inferred depends on the level of additional knowledge (ie, “glue”) one adds to the original data
- More complex ontologies: more inference possibilities, but more complex reasoning procedures
- At present, W3C has defined a set of ontology languages (and is working on rules)
- An application may choose the complexity it wants



"One has to learn formal logic, knowledge representation techniques, description logic, etc"

Not really…

- Yes, the detailed semantics of RDFS, OWL Lite, etc, are based on knowledge representation algorithms
  - *OWL-DL stands for "OWL Description Logic"; it is an embodiment of a Description Logic*
- …but most users just have to *use* these
- It is just like SQL: the formal semantics is *very* complex, but 95% of the SQL users have never even looked at it!
- *Developing* and ontology may require more knowledge, but that is for a small percentage of users (and there are authoring tools to hide the details)

Where do the data and ontologies come from?

(Should we really expect the author to type in all this data?)

Pure RDF data: not always a solution…

- Creating large scale RDF data with an editor is possible, but does not really scale…
    - *although it may be o.k. for small things like the "glue" in our example*
- Even if it is around: adding RDF to, say, XHTML, is not always easy
    - *there are number of disagreeable technical problems with, eg, validation*
    - *the only "clean" approach today is to link it via a* **`meta`** *header element*

Data may be around already…

- Part of the (meta)data information is present in tools … but thrown away at output
  - *e.g., a business chart can be generated by a tool…*
  - *…it "knows" the structure, the classification, etc. of the chart, but, usually, this information is lost*
- storing it in web data would be easy!
- "SW-aware" tools are around (even if you do not know it…), though more would be good:
  - *Photoshop CS stores metadata in RDF in, say, jpg files (referred to as XMP)*
  - *RSS 1.0 feeds are generated by (almost) all blogging systems (a huge amount of RDF data!)*
  - *…*

Data may be extracted (a.k.a. “scraped”)

- Different tools, services, etc, come around every day:
  - *get RDF data associated with images, for example:*
    - service to get RDF from flickr images (see example)
    - service to get RDF from XMP (see example)
  - *XSLT scripts to retrieve microformat based information from XHTML files*
  - *scripts to convert spreadsheets to RDF*
  - *etc*
- Most of these tools are still individual “hacks”, but show a general tendency

Formalizing the scraper approach: GRDDL

- GRDDL *formalizes* the scraper approach. For example:

```
<html xmlns="http://www.w3.org/1999/">
  <head profile="http://www.w3.org/2003/g/data-view">
    <title>Some Document</title>
    <link rel="transformation" href="http:…/dc-extract.xsl"/>
    <meta name="DC.Subject" content="Some subject"/>
    ...
  </head>
  ...
  <span class="date">2006-01-02</span>
  ...
</html>
```

- yields, by running the file through `dc-extract.xsl`

```
<rdf:Description rdf:about="…">
  <dc:subject>Some subject</dc:subject>
  <dc:date>2006-01-02</dc:date>
</rdf:Description>
```

GRDDL (cont)

- Somebody has to provide `dc-extract.xsl` and use its conventions (making use of the corresponding meta-s, class id-s, etc…)
- … but, by using the `profile` attribute, a client is instructed to find and run the transformation processor automatically
- A “bridge” to “microformats”
- A W3C Working Group has just started, with a recommendation planned in the 1st Quarter of 2007

Another Future Solution: RDFa

- RDFa (formerly known as RDF/A) extends XHTML by:
  - *extending the* `link` *and* `meta` *elements (e.g., meta elements may have children, thereby adding more complex data; usable throughout the body, too)*
  - *defining general attributes to add metadata to any elements (a bit like the* `class` *in microformats, but via dedicated properties)*
- It is very similar to microformats, but with more rigor:
  - *it is a general framework (instead of an “agreement” on the meaning of, say, a* `class` *attribute value)*
  - *terminologies can be mixed more easily*
- The W3C Working Group on SW Deployment has this on its charter

RDFa example

- For example

```
<div about="http://uri.to.newsitem">
  <span property="dc:date">March 23, 2004</span>
  <span property="dc:title">Rollers hit casino for £1.3m</span>
  By <span property="dc:creator">Steve Bird</span>. See
  <a href="http://www.a.b.c/d.avi" rel="dcmtype:MovingImage">
  also video footage</a>…
</div>
```

- yields, by running the file through a processor:

```
<http://uri.to.newsitem>
  dc:date             "March 23, 2004";
  dc:title            "Rollers hit casino for £1.3m;
  dc:creator          "Steve Bird";
  dcmtype:MovingImage <http://www.a.b.c/d.avi>.
```

Linking to SQL

- A huge amount of data in Relational Databases
- Although tools exist, it is not feasible to *convert* that data into RDF
- Instead: SQL ⇋ RDF “bridges” are being developed:
  - *a query to RDF data is transformed into SQL on-the-fly*
  - *the modalities are governed by small, local ontologies or rules*
- An active area of development!

And for Ontologies?

- The hard work is to *create* the ontologies in general
  - *requires a good knowledge of the area to be described*
  - *some communities have good expertise already (e.g., librarians)*
  - *OWL is just a tool to formalize ontologies*
- Large scale ontologies are often developed in a community process
  - *leading to versioning issues, too*
  - *OWL includes predicates for versioning, deprecation, "same-ness", …*
- There is also R&D in *generating* them from a corpus of data
  - *still mostly a research subject*
- Sharing ontologies may be vital in the process

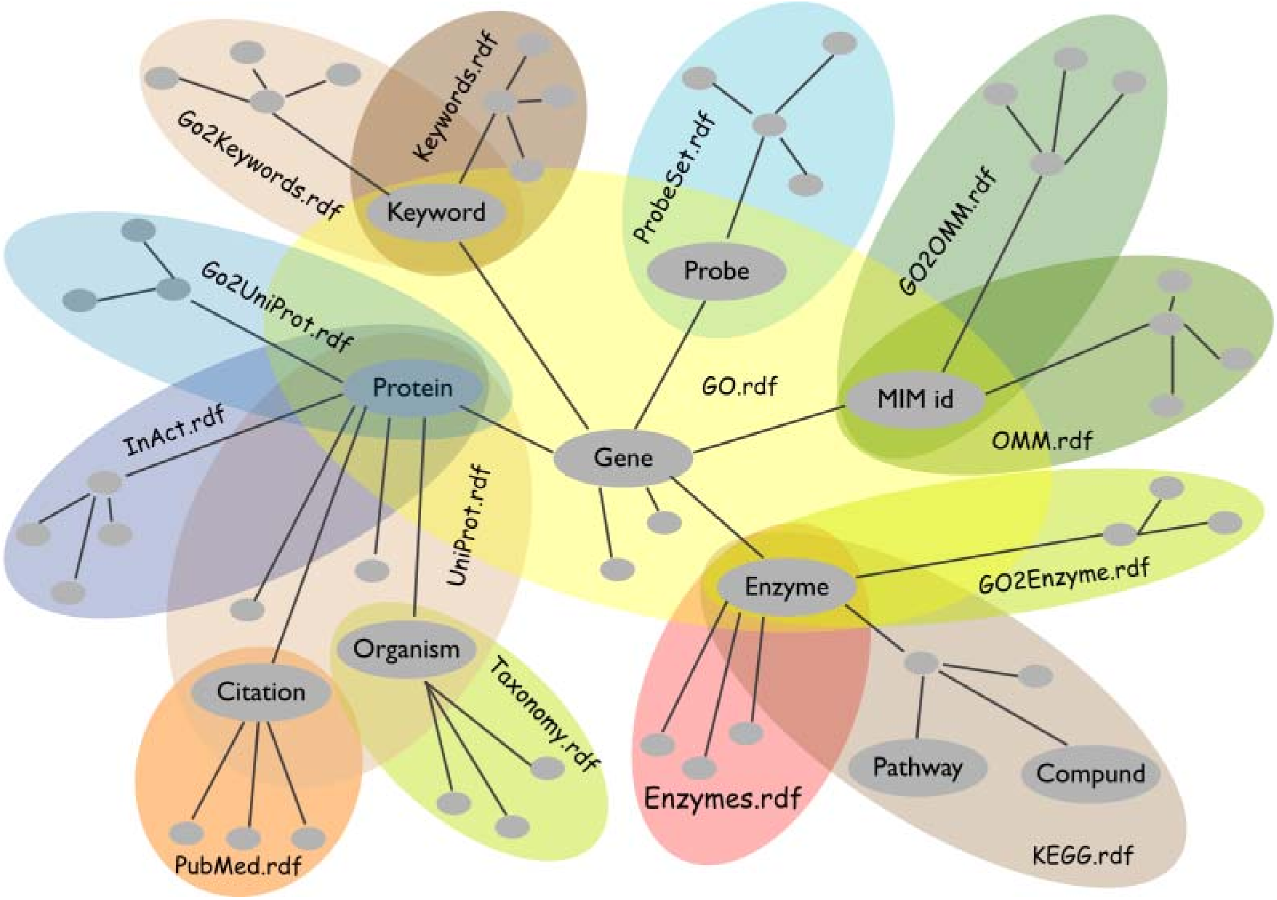
There are already ontologies around…

- Lots of ontologies registered at Schemaweb
- DAML ontology library has several hundreds of ontologies
- Ontologies are being developed by various communities:
  - *medical domain (e.g., the US Cancer Institute's Cancer Ontology, the Gene Ontology, the BioPax Molecular Pathway Ontology, …)*
  - *cultural heritage domain (e.g., CIDOC reference model and ontology)*
  - *OWL representation of (English) Wordnet*
  - *eBusiness ontology for products and services: eClassOwl*
  - *…*
- *Use existing ontologies when you can!*

“Core” vocabularies

- A number of public “core” vocabularies evolve to be used by applications, e.g.:
  - *SKOS Core: about knowledge systems*
  - *Dublin Core: about information resources, digital libraries, with extensions for rights, permissions, digital right management*
  - *FOAF: about people and their organizations*
  - *DOAP: on the descriptions of software projects*
  - *MusicBrainz: on the description of CDs, music tracks, …*
  - *SIOC: Semantically-Interlinked Online Communities*
  - *…*

A mix of ontologies (a life science example)…



How do I extract triplets from and RDF Graph? Ie: how do I query an RDF Graph?

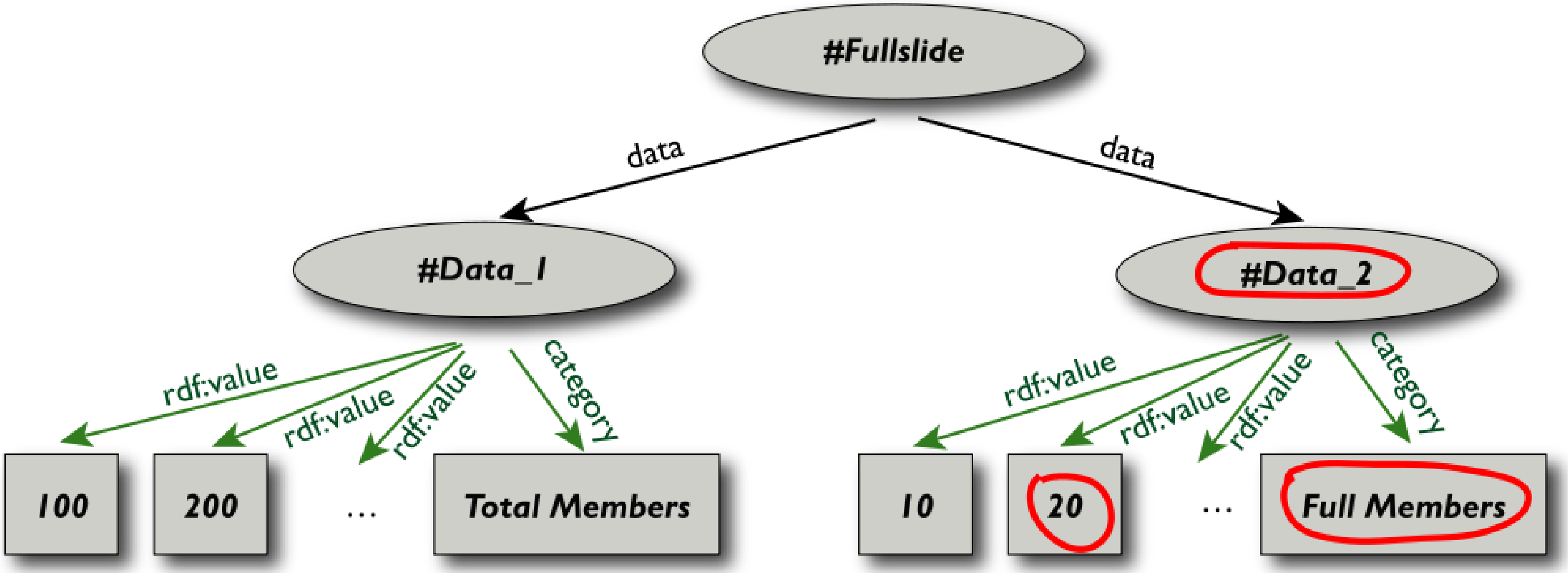
Querying RDF graphs

- The fundamental idea: use *graph patterns* to define subgraphs:
  - *a pattern contains unbound symbols*
  - *by binding the symbols, subgraphs of the RDF graph may be matched*
  - *if there is such a match, the query returns the bound resources or a subgraph*
- This is the how SPARQL (Query Language for RDF) is defined

Simple SPARQL Example

```
SELECT ?cat ?val # note: not ?x!
WHERE { ?x rdf:value ?val. ?x category ?cat }
```

- Returns: `[["Total Members",100],["Total Members",200],…,["Full Members",10],…]`



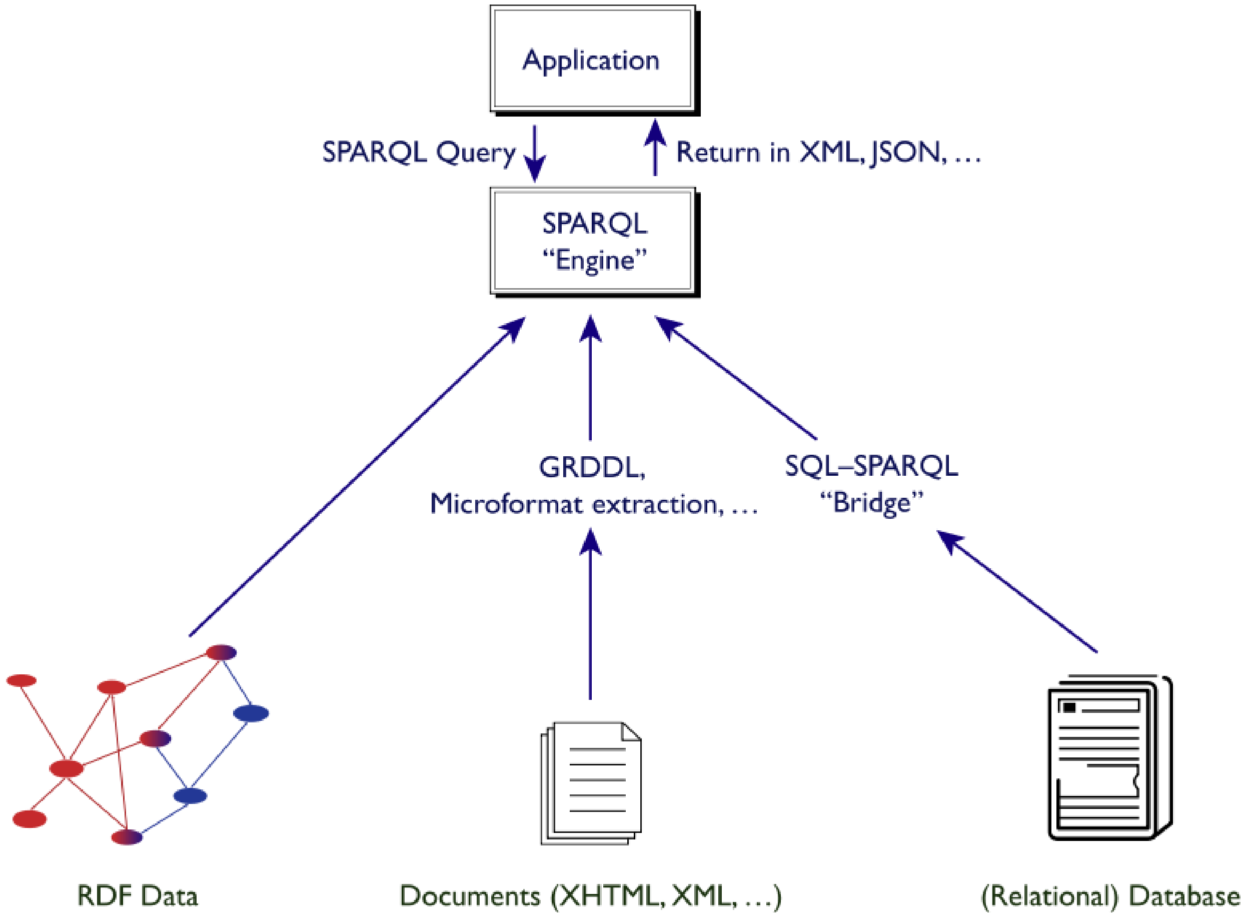
Other SPARQL features

- Define *optional* patterns
- Limit the number of returned results; remove duplicates, sort them,…
- Add functional constraints to pattern matching
- Return a full *subgraph* (instead of a list of bound variables)
- Use datatypes and/or language tags when matching a pattern
- SPARQL is not yet finalized, but will become a Recommendation (hopefully) in 2nd Quarter of 2007
  - *but there are a number of implementations already!*

SPARQL usage in practice

- *Locally*, i.e., bound to a programming environment like RDFLib or Jena
  - *details are language dependent*
- *Remotely*, e.g., over the network or into a database
  - *very important usage: a growing number of RDF depositories…*
  - *separate documents define the protocol and the result format*
    - SPARQL Protocol for RDF
    - SPARQL Results XML Format
    - there is also a JSON binding (soon a W3C note…)
- An application pattern evolves: use (XHTML) forms to create a SPARQL Query to a database and display the result in XHTML
  - *there are a number of application experiments, demos, etc.,*

SPARQL as a federating tool



Isn't This Research Only?

(or: does this have *any* industrial relevance whatsoever?)

Not any more…

- *Lots* of tools are available. Are listed on W3C's wiki:
  - *RDF programming environment for 14+ languages, including C, C++, Python, Java, Javascript, Ruby, PHP,… (no Cobol or Ada yet ☹!)*
  - *13+ Triple Stores, ie, database systems to store (sometimes huge!) datasets*
  - *a number programming environments (in Java, Prolog, …) include OWL reasoners*
  - *there are also stand-alone reasoners (downloadable or on the Web)*
  - *etc*
- Some of the tools are Open Source, some are not; some are very mature, some are not 😉: *it is the usual picture of software tools*, nothing special any more!
- *Anybody can start developing RDF-based applications today*

Not any more… (cont)

- SW has indeed a strong foundation in research results
- But remember:
  - *(1) the Web was born at CERN…*
  - *(2) …was first picked up by high energy physicists…*
  - *(3) …then by academia at large…*
  - *(4) …then by small businesses and start-ups…*
  - *(5) "big business" came only later!*
- network effect kicked in early…
- Semantic Web is now at #4, and moving to #5!

Some RDF deployment areas

	Library metadata	Defence	Life sciences
Problem to solve?	single-domain integration	yes, serious data integration needs	yes, connections among genetics, proteomics, clinical trials, regulatory, …
Willingness to adopt?	yes: OCLC push and Dublin Core initiative	yes: funded early DAML (OWL) work	yes: intellectual level high, much modeling done already.
Motivation	light	strong	very strong
Links to	other library data	phone calls records, etc	chemistry, regulatory, medical, etc
Showcase?	limited	not at all	yes, model for other industries.

Some RDF deployment areas (cont)

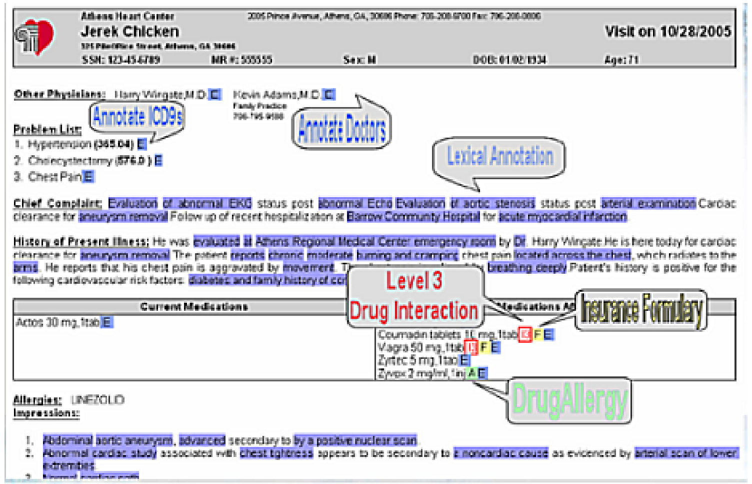
- These are just examples
- Others are coming to the fore: eGovernment, energy sector (oil industry), financial services, …
- Health care and life science sector is now very active
  - *also at W3C, in the form of an Interest Group*

The “corporate” landscape is moving

- See, for example, the Semantic Technology Conference series
  - *not a scientific conference, but commercial people making real money!*
  - *speakers in 2006: from IBM, Cisco, BellSouth, GE, Walt Disney, Nokia, Oracle, …*
  - *not all referring to Semantic Web (eg, RDF, OWL, …) but semantics in general*
  - *but they might come around!*
- Major companies offer (or will offer) Semantic Web tools or systems using Semantic Web: Adobe, Oracle, IBM, HP, Software AG, webMethods, Northrop Gruman, Altova, …
- “Corporate Semantic Web” listed as major technology by Gartner in 2006

Applications are not always very complex…

- Eg: simple semantic annotations of patients’ data greatly enhances communications among doctors
- What is needed: some simple ontologies, an RDFa/microformat type editing environment
- Simple but powerful!



Athens Heart Center 2005 Prince Avenue, Athens, GA, 30606 Phone: 706-208-0700 Fax: 706-208-0006

Jerek Chicken

Visit on 10/28/2005

125 PineOffice Street, Athens, GA 30606

SSN: 123-45-6789 MR #: 555555 Sex: M DOB: 01/02/1934 Age: 71

Other Physicians: Harry Wingate,M.D. E Kevin Adams,M.D. E
Family Practice
706-795-9588

Problem List:

1. Hypertension **(365.04)** E
2. Cholecystectomy **(576.0)** E
3. Chest Pain E

Chief Complaint: Evaluation of abnormal EKG status post abnormal Echo Evaluation of aortic stenosis status post arterial examination Cardiac clearance for aneurysm removal Follow up of recent hospitalization at Barrow Community Hospital for acute myocardial infarction

History of Present Illness: He was evaluated at Athens Regional Medical Center emergency room by Dr. Harry Wingate He is here today for cardiac clearance for aneurysm removal The patient reports chronic moderate burning and cramping chest pain located across the chest, which radiates to the arms. He reports that his chest pain is aggravated by movement. T[illegible] breathing deeply Patent's history is positive for the following cardiovascular risk factors: diabetes and family history of co[illegible]

Current Medications	Medications A[illegible]
Actos 30 mg, 1tab E	Coumadin tablets 10 mg, 1tab I F E Viagra 50 mg, 1tab I F E Zyrtec 5 mg, 1tab E Zyvox 2 mg/ml, 1inj A E

Allergies: LINEZOLID

Impressions:

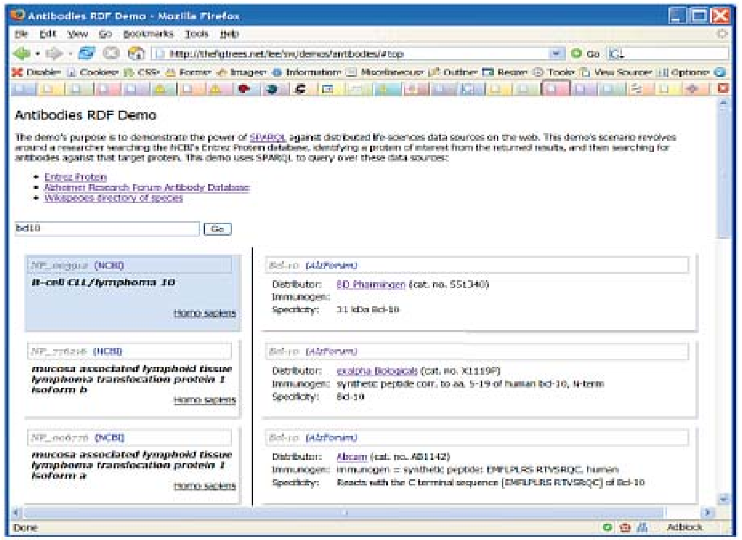
1. Abdominal aortic aneurysm, advanced secondary to by a positive nuclear scan
2. Abnormal cardiac study associated with chest tightness appears to be secondary to a noncardiac cause as evidenced by arterial scan of lower extremities
3. Normal cardiac cath

Data integration

- Data integration comes to the fore as one of *the* SW Application areas
- Very important for large application areas (life sciences, energy sector, eGovernment, financial institutions), as well as everyday applications (eg, reconciliation of calendar data)
- Life sciences example:
  - *data in different labs…*
  - *data aimed at scientists, managers, clinical trial participants…*
  - *large scale public ontologies (genes, proteins, antibodies, …)*
  - *different formats (databases, spreadsheets, XML data, XHTML pages)*
  - *etc*

Example: antibodies demo

- Scenario: find the known antibodies for a protein in a specific species
- Combine ("scrape"…) three different data sources
- Use SPARQL as an integration tool (see also demo online)



Antibodies RDF Demo - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://thefigtrees.net/lee/sw/demos/antibodies/#top
Go
Disable Cookies CSS Forms Images Information Miscellaneous Outline Resize Tools View Source Options
Done
Adblock

Antibodies RDF Demo

The demo's purpose is to demonstrate the power of SPARQL against distributed life-sciences data sources on the web. This demo's scenario revolves around a researcher searching the NCBI's Entrez Protein database, identifying a protein of interest from the returned results, and then searching for antibodies against that target protein. This demo uses SPARQL to query over these data sources:

- Entrez Protein
- Alzheimer Research Forum Antibody Database
- Wikispecies directory of species

bcl10 Go

NP_003912 (NCBI)
B-cell CLL/lymphoma 10
Homo sapiens

NP_776216 (NCBI)
mucosa associated lymphoid tissue lymphoma translocation protein 1 isoform b
Homo sapiens

NP_006776 (NCBI)
mucosa associated lymphoid tissue lymphoma translocation protein 1 isoform a
Homo sapiens

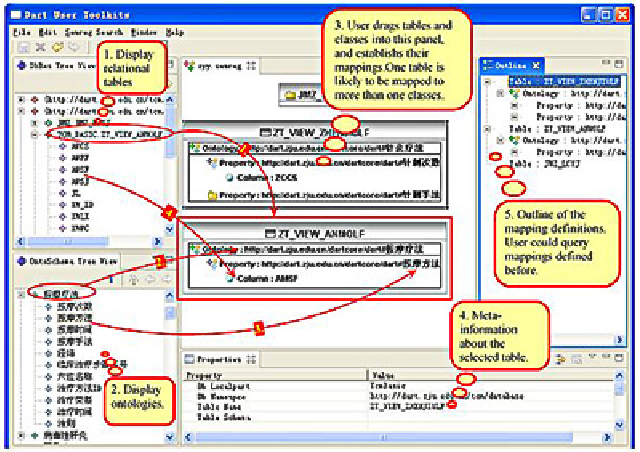
Bcl-10 (AlzForum)
Distributor: BD Pharmingen (cat. no. 551340)
Immunogen:
Specificity: 31 kDa Bcl-10

Bcl-10 (AlzForum)
Distributor: exalpha Biologicals (cat. no. X1119P)
Immunogen: synthetic peptide corr. to aa. 5-19 of human bcl-10, N-term
Specificity: Bcl-10

Bcl-10 (AlzForum)
Distributor: Abcam (cat. no. AB1142)
Immunogen: immunogen = synthetic peptide: EMFLPLRS RTVSRQC, human
Specificity: Reacts with the C terminal sequence [EMFLPLRS RTVSRQC] of Bcl-10

There has been lots of R&D

- Boeing, MITRE Corp., Elsevier, EU Projects like Sculpteur and Artiste, national projects like MuseoSuomi, DartGrid, …
- Developments are under way at various places in the area



1. Display relational tables

2. Display ontologies.

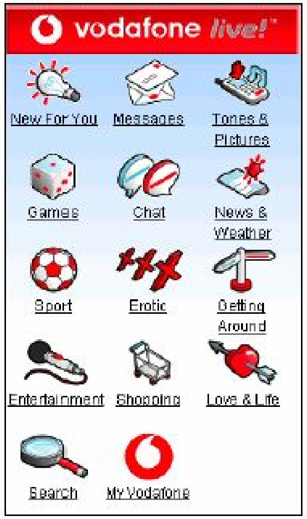
3. User drags tables and classes into this panel, and establishs their mappings.One table is likely to be mapped to more than one classes.

4. Meta-information about the selected table.

5. Outline of the mapping definitions. User could query mappings defined before.

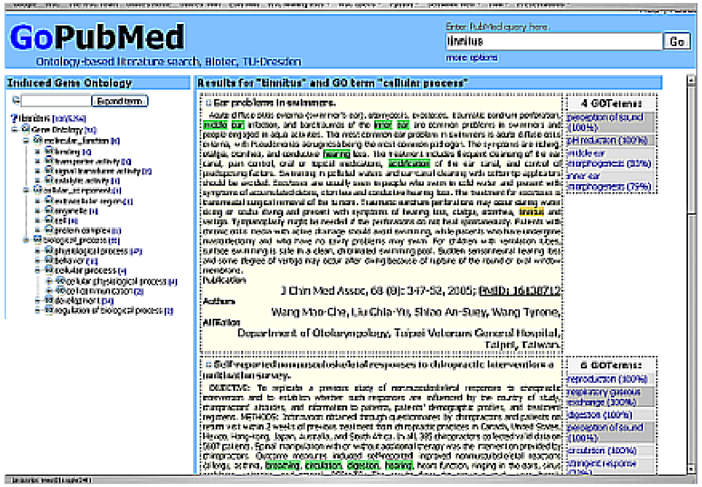
Portals

- Vodafone's Live Mobile Portal
  - *search application (e.g. ringtone, game, picture) using RDF*
    - page views per download decreased 50%
    - ringtone up 20% in 2 months
- Sun’s SwordFish: public queries for support, handbooks, etc, go through an internal RDF engine for White Paper Collections and System Handbook collections
- Nokia has a somewhat similar support portal
- Harper’s Online magazine links items together via an internal ontology
- See also Opera’s presentation later today…



Improved Search via Ontology: GoPubMed

- Improved search on top of pubmed.org
  - *search results are ranked using the specialized ontologies*
  - *extra search terms are generated and terms are highlighted*
- Importance of *domain specific ontologies* for search improvement



GoPubMed

Ontology-based literature search, Biotec, TU-Dresden

Enter PubMed query here.

tinnitus

more options

Induced Gene Ontology

Expand term

- tinnitus [illegible]
  - Gene Ontology [illegible]
    - molecular_function [illegible]
      - binding [illegible]
      - transporter activity [illegible]
      - signal transducer activity [illegible]
      - catalytic activity [illegible]
    - cellular_component [illegible]
      - extracellular region [illegible]
      - organelle [illegible]
      - cell [illegible]
      - protein complex [illegible]
    - biological_process [illegible]
      - physiological process [illegible]
      - behavior [illegible]
      - cellular process [illegible]
        - cellular physiological process [illegible]
        - cell communication [illegible]
      - development [illegible]
      - regulation of biological process [illegible]

Results for "tinnitus" and GO term "cellular process"

Ear problems in swimmers.

Acute diffuse otitis externa (swimmer's ear), otomycosis, exostoses, traumatic eardrum perforation, middle ear infection, and barotraumas of the inner ear are common problems in swimmers and people engaged in aqua activities. [illegible] hearing loss. [illegible] acidification of the ear canal, and control of [illegible] tinnitus and vertigo. [illegible] membrane.

Publication

J Chin Med Assoc, 68 (8): 347-52, 2005; PMID: 16138712

Authors

Wang Mao-Che, Liu Chia-Yu, Shiao An-Suey, Wang Tyrone,

Affiliation

Department of Otolaryngology, Taipei Veterans General Hospital, Taipei, Taiwan.

4 GOTerms:

perception of sound (100%)
pH reduction (100%)
middle ear morphogenesis (83%)
inner ear morphogenesis (75%)

Self-reported musculoskeletal responses to chiropractic intervention: a multination survey.

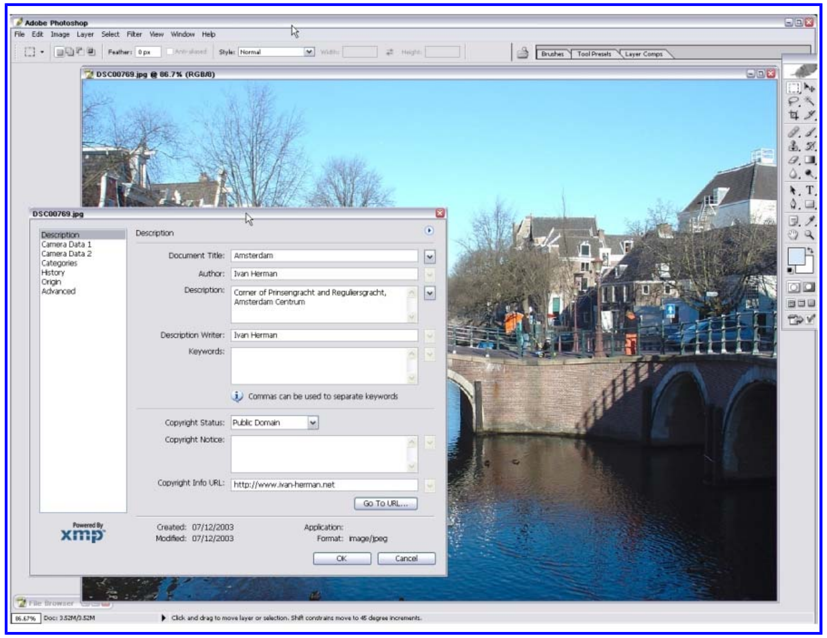
OBJECTIVE: To replicate a previous study of musculoskeletal responses to chiropractic intervention and to establish whether such responses are influenced by the country of study, chiropractors' attitudes, and information to patients, patients' demographic profiles, and treatment regimens. METHODS: Information obtained through questionnaires by chiropractors and patients on return visit within 2 weeks of previous treatment from chiropractic practices in Canada, United States, Mexico, Hong-Kong, Japan, Australia, and South Africa. In all, 385 chiropractors collected valid data on [illegible] breathing, circulation, digestion, hearing, [illegible]

6 GOTerms:

reproduction (100%)
respiratory gaseous exchange (100%)
digestion (100%)
perception of sound (100%)
circulation (100%)
[illegible] response [illegible]

Adobe's XMP

- Adobe’s (public) tool to add RDF-based metadata to *most* of their file formats
  - *supported in Adobe Creative Suite*
  - *support from 30+ major asset management vendors, with separate XMP conferences; will be used in Windows Vista*



Description

Document Title: Amsterdam

Author: Ivan Herman

Description: Corner of Prinsengracht and Reguliersgracht, Amsterdam Centrum

Description Writer: Ivan Herman

Keywords:

Commas can be used to separate keywords

Copyright Status: Public Domain

Copyright Notice:

Copyright Info URL: http://www.ivan-herman.net

Created: 07/12/2003

Modified: 07/12/2003

Application:

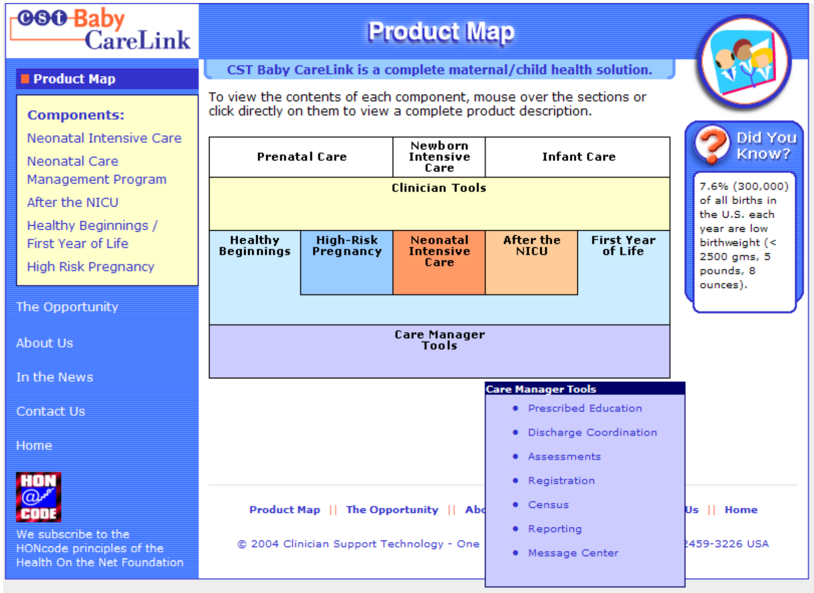
Format: image/jpeg

86.67% Doc: 3.52M/3.52M

Click and drag to move layer or selection. Shift constrains move to 45 degree increments.

Baby CareLink

- Centre of information for the treatment of premature babies
- Provides an OWL service *as a Web Service*
  - *combines disparate vocabularies like medical, insurance, etc*
  - *users can add new entries to ontologies*
  - *complex questions can be asked through the service*



CST Baby CareLink is a complete maternal/child health solution.

To view the contents of each component, mouse over the sections or click directly on them to view a complete product description.

7.6% (300,000) of all births in the U.S. each year are low birthweight (< 2500 gms, 5 pounds, 8 ounces).

We subscribe to the HONcode principles of the Health On the Net Foundation

© 2004 Clinician Support Technology - One
459-3226 USA

Summary

- The Semantic Web is not as complex as people believe
- The Semantic Web does not require huge investments before seeing its value
- The Semantic Web is not only for geeks…

Thank you for your attention!